

Assessing joint spatial autocorrelations between mortality rates due to cardiovascular conditions in South Africa

Timotheus B. Darikwa,¹ Samuel Manda,^{2,3} Maseka Lesaoana¹

¹Department of Statistics and Operations Research, University of Limpopo, Polokwane; ²Biostatistics Research Unit, South African Medical Research Council, Pretoria; ³Department of Statistics, University of Pretoria, Hatfield, South Africa

Abstract

South Africa is experiencing an increasing burden of non-communicable diseases (NCDs). There is evidence of co-morbid-

ity of several NCDs at small geographical areas in the country. However, the extent to which this applies to joint spatial autocorrelations of NCDs is not known. The objective of this study was to derive and quantify multivariate spatial autocorrelations for NCD-related mortality in South Africa. The study used mortality attributable to cerebrovascular, ischaemic heart failure and hypertension captured by the country's Department of Home Affairs for the years 2001, 2007 and 2011. Both univariate and pairwise spatial clustering measures were derived using observed, empirical Bayes smoothed and age-adjusted standardised mortality rates. Cerebrovascular and ischaemic heart co-clustering was significant for the years 2001 and 2011. Cerebrovascular and hypertension co-clustering was significant for the years 2007 and 2011, while hypertension and ischaemic heart co-clustering was significant for the year 2011. Co-clusters of cerebrovascular-ischaemic heart disease are the most profound and located in the south-western part of the country. It was successfully demonstrated that bivariate spatial autocorrelations can be derived for spatially dependent mortality rates as exemplified by mortality rates attributed to three cardiovascular conditions. The identified co-clusters of spatially dependent health outcomes may be targeted for an integrated intervention and monitoring programme.

Correspondence: Timotheus B. Darikwa, Department of Statistics and Operations Research, University of Limpopo, Private Bag X1106, Sovenga, 0727, South Africa.
Tel. +27.015.2683674 - Fax: +27.015.2683487.
E-mail: Timotheus.Darikwa@ul.ac.za

Acknowledgments: the authors would like to thank Statistics South Africa for providing the data used for analysis.

Key words: Bivariate spatial autocorrelation; Indirect standardised mortality rate; Cardiovascular mortality; Empirical Bayes smoothing; South Africa.

See online Appendix for an additional table.

Contributions: TBD reviewed the relevant literature, performed data management and statistical analysis and wrote all drafts of manuscript; SM conceived the ideas, reviewed the statistical analysis and helped with the revision of the manuscripts; LM helped with the revision of the manuscript and statistics. All authors have read and approved the final manuscript.

Conflict of interest: the authors declare no potential conflict of interest.

Funding: the study was supported by the following grants or organisations at different stages: SAMRC-Biostatistics Capacity Development, no. 57042; Teaching development grant national collaborative project, no. APP-TDG-088; DST-NRF Centre of Excellence in Mathematical and Statistical Sciences (CoE-MASS); VLIR-UOS. Opinions expressed and conclusions arrived at are those of the authors and are not necessarily to be attributed to UL, SAMRC, CoE-MASS or VLIR-UOS.

Received for publication: 29 April 2019,
Revision received: 18 October 2019.
Accepted for publication: 18 October 2019.

© Copyright: the Author(s), 2019
Licensee PAGEPress, Italy
Geospatial Health 2019; 14:784
doi:10.4081/gh.2019.784

This article is distributed under the terms of the Creative Commons Attribution Noncommercial License (CC BY-NC 4.0) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Introduction

Global mortality attributed to non-communicable diseases (NCDs) has become so grave in recent times that NCDs have been included as one of the 17 sustainable developmental goals (SDGs) targets where premature mortality (before 70 years of age) attributed to the diseases have to be reduced by one third by the year 2030 according to the World Health Organization (WHO) (WHO, 2015). NCDs killed 41 million (71%) of the 57 million people who died in the world in 2016 (WHO, 2018). About 44% (17.9 million) of the NCD deaths were attributed to cardiovascular diseases (CVDs) making them the leading NCD killers in the world (WHO, 2018). A review of progress towards the 2030 target of a one-third reduction suggest the target is going to be missed (WHO, 2018). Thus, individual countries need to increase efforts if the target is to be met. South Africa is one country where NCDs are currently the leading burden of diseases (Groenewald *et al.*, 2014). Cardiovascular diseases (CVDs) contribute the highest proportion (48%) of South Africa's NCD deaths (Bradshaw *et al.*, 2006). Thus containing CVDs will contribute significantly towards reaching the SDG target for South Africa. CVDs are known to cluster geographically, depending on levels of deprivation (Ford and Highfield, 2016). In South Africa, the previously disadvantaged or deprived communities are exhibiting higher prevalence levels of metabolic syndrome (a collection of risk fac-



tors for CVDs and diabetes; Alberti and Zimmet, 2005) than the advantaged communities, putting areas populated by a majority of Blacks at high risk of mortality due to CVDs. This is attributed to, among other factors, a nutritional transition taking place in the country. It is of great importance to understand, and monitor the changing spatial patterns of CVDs and identify cluster areas of high mortality risk of NCDs if the SDG target is to be met. It is the objective of this study to assess pairwise common spatial co-clustering (joint clusters) of the three major CVD-related mortality in South Africa.

Investigation of the spatial variation and clustering of mortality in South Africa is not new. Descriptive risk maps have been used to describe the geographic variation of NCD mortality (Bradshaw *et al.*, 1995, 2006; Groenewald *et al.*, 2014) while univariate spatial autocorrelation measures have been used to determine the presence of spatial heterogeneity or variation in an area and to detect clusters of human immunodeficiency virus/acquired immune deficiency syndrome (HIV/AIDS) mortality (Tanser *et al.*, 2009), infant mortality (Sartorius *et al.*, 2011) and all-cause mortality (Sartorius *et al.*, 2010). Univariate global indicators of spatial autocorrelation (GISA) are used to detect spatial heterogeneity or variation of cause-specific mortality for an area. They test the extent to which neighbours are similar or different in the region of study. They can be used to confirm if there is any form of clustering but they do not reveal actual clusters (Waller and Gotway, 2004). Actual clusters are detected using local indices of spatial autocorrelation (LISA). Using univariate LISA to identify clusters entails investigating if high mortality risk of a single disease in a given area extend to neighbouring areas. Areas of high mortality risk that extend to nearby areas form a cluster known as hotspots, while areas of low mortality risk that extend to nearby areas form coldspots for the disease in question.

Joint mapping of multiple disease outcomes has also been done in South Africa using the shared-spatial component method to establish ecological associations between HIV/AIDS and syphilis (Manda *et al.*, 2012) as well as for some CVDs (Kandala *et al.*, 2014). This method only shows that multiple diseases spatially co-exist but does not explore the influence of the existence of a disease in an area on the spread of the disease in adjacent areas and/or the influence of one disease in an area on the spread of other diseases in neighbouring areas. Sometimes, when dealing with interrelated diseases like CVDs it is important to determine how they influence each other spatially. Univariate spatial autocorrelations cannot do this, and neither can the spatial component method. However, multivariate spatial autocorrelation techniques can give more insight into the spatial dependency of two or more interrelated disease outcomes. This study considers the application of bivariate spatial autocorrelation measures by Lee (2001), Anselin *et al.* (2002) and Dray *et al.* (2008) in order to detect high mortality risk pairwise co-clustering of three interrelated CVDs around local municipal neighbourhoods of South Africa.

Bivariate spatial dependence between mortality rates due to three leading cardiovascular conditions of cerebrovascular, hypertension and ischaemic heart failure was determined using South African data for the years 2001, 2007 and 2011 with the local municipality as the unit of measurement. Both raw rates and smoothed rates (Clayton and Kaldor, 1987) were used and results compared. The later rates would be preferable as a number of municipalities had sparse data in both counts and populations that may result in rates instability when using crude rates (Marshall, 1991). This study set out to use bivariate spatial autocorrelation

measures to detect pairwise co-clustering in two ways. Firstly, in a given period, to test if the existence of high risk of mortality of a disease in an area extend to nearby areas for a related disease. The hypothesis being tested here is that interrelated diseases co-cluster. Identifying co-clusters of CVDs is important if a unified approach is to be employed in the prioritisation, prevention of the spread, diagnosis and cure of the related diseases. Secondly, it was used to test if the spatial patterns of a disease differ between two time points. It is important to do this second test in order to establish if the spatial dynamics of a disease are changing with time. If the spatial dynamics of a disease are stable it is much easier to predict the spatial patterns of the disease over time for planning and monitoring purposes.

Materials and Methods

Data

Causes of death (COD) data from South Africa's vital registration system were used in this study. The data are collected using the death notification forms (DNFs). Medical personnel and other approved certifying authorities are allowed to complete the DNFs. Information collected is kept by the South African Department of Home Affairs, who in turn allow Statistics South Africa (Stats SA) to collate the COD data for onward distribution to the public. Stats SA uses revision number ten of the International Statistical Classification of Diseases and Related Health Problems (ICD-10) to code and classify the COD data as stipulated by the WHO (WHO, 2004).

This paper only considers ICD-10 defined broad groups of COD data due to three leading CVDs causing mortality in South Africa. These are cerebrovascular heart diseases (CVAs), hypertensive heart diseases (HHDs) and ischaemic heart diseases (IHDs) for the years 2001, 2007 and 2011. Table 1 shows the total deaths due to IHD increased the most for the period under review from 10769 in 2001 to 15609 deaths in 2011. An increase of 45%. CVA deaths increased by 14% from 22 590 to 25983, while IHD increased by only 2% (11779 to 12023) over the same period. Generally, the national mean mortality rates were constant over the period for the three CVAs.

Joubert *et al.* (2013) outlined the data quality issues associated with DNF data. Problems associated with these data include, among others, garbage codes, misclassification and incompleteness of death registration. Adjustments were made to these data to minimise bias that may be attributed to misclassification and incompleteness of death registration. Firstly, the death data per municipality were adjusted for incompleteness of death registration using provincial percentages that were derived by Pillay-Van Wyk *et al.* (2014). The assumption we make is that completeness of death registration is the same throughout the province for each cause of death. The completeness of death registration for the age group 5-14 years was derived through interpolation for each province by assuming a linear relationship between the ages 1-4 and 15+ as advised by Pillay-Van Wyk *et al.* (2014). Appendix Table A1 shows the completeness of death registration values used for adjusting for underreporting of deaths in each province for the given age groups.

Misclassification of deaths is another common problem with South African cause of death data (Birnbbaum *et al.*, 2011; Joubert

et al., 2013; Groenewald *et al.*, 2014; Pillay-Van Wyk *et al.*, 2014). The extent of misclassification of cardiovascular mortality is unknown in South Africa. In order to correct the rate of cardiovascular mortality, the vital registration data for the years 2001, 2007 and 2011 were pooled to get aggregate distributions of deaths by cause and age for the whole of South Africa. These data were then used to derive the national cause-specific death rates by age. Similarly, cause-specific death rates by age for the years 2001, 2007 and 2011 were computed for each of South Africa's 213 municipalities.

Assuming that national specific-mortality rates by age groups are more accurate at national level than at local level (Joubert *et al.*, 2013), municipality-level standardised mortality rates (SMRs) were computed using indirect standardisation. Using global specific mortality rates by age groups as in Birnbaum *et al.* (2011) produced similar municipality-level SMRs. Thus in this paper, three mortality rates were used for analysis, namely, raw mortality rates, SMRs and empirically Bayes (EB) smoothed rates (Marshall, 1991; Leyland and Davies, 2005).

A descriptive summary of the original data, raw rates, EB rates and SMRs used is given in Table 1. The effects of EB smoothing and indirect standardisation of the raw mortality rates differ by diseases and year. In the case of EB rates, maximum values are lower than those of the raw rates in all instances. Further investigations revealed that the municipalities with the smallest populations are

also the ones with the highest (as well as smallest) mortality rates. The raw rates are sensitive to small population counts resulting in instability. This is a well-documented problem when using raw mortality rates and the Empirically Bayes smoothed rates alleviate these problems inherent in most crude or raw rates (Marshall, 1991; Leyland and Davies, 2005).

Spatial statistics methods

The COD data are recorded at local municipal level prompting municipalities to be considered as fixed and countable areal units. Subsequently, areal data spatial autocorrelation measures were considered for analysing the data. Univariate and bivariate spatial autocorrelations are presented in this section.

Global indexes of spatial autocorrelation

GISA are used to determine the extent to which neighbours are similar in the region of study. They can be used to confirm if there is any form of clustering but they do not reveal actual clusters (Waller and Gotway, 2004). There are a number of Global tests such as the quadrat method, the nearest neighbour method, Geary's *C* and the global Moran's *I* test. The most popular of the Global measures of spatial autocorrelation is the global Moran's *I* statistic and is the one we discuss here.

Table 1. Distribution of number and death rates across municipalities by year, South Africa.

Variable	Rate type	Mean	Min	Max	Total number of deaths
CVA01	Raw Rate	56	0	274	22,590
	EB Rate	56	0	231	
	SMR	64	0	245	
CVA07	Raw Rate	57	0	186	25,391
	EB Rate	58	3	183	
	SMR	59	0	202	
CVA11	Raw Rate	59	0	318	25,893
	EB Rate	58	5	310	
	SMR	57	0	256	
IHD01	Raw Rate	29	0	270	11,779
	EB Rate	29	1	254	
	SMR	31	0	236	
IHD07	Raw Rate	26	0	132	12,489
	EB Rate	26	0	119	
	SMR	25	0	99	
IHD11	Raw Rate	25	0	154	12,023
	EB Rate	34	3	137	
	SMR	23	0	128	
HHD01	Raw Rate	24	0	93	10,769
	EB Rate	25	2	90	
	SMR	29	0	98	
HHD07	Raw Rate	29	0	110	13,379
	EB Rate	30	2	106	
	SMR	30	0	108	
HHD11	Raw Rate	34	0	146	15,609
	EB Rate	25	1	130	
	SMR	33	0	116	

CVA, cerebrovascular heart disease; IHD, ischaemic heart disease; HHD, hypertensive heart disease; EB, empirically Bayes; SMR, standardised mortality rates.



Global Moran's I

The global univariate Moran's *I* statistic measures the extent of the linear relation between the observed geo-referenced data $x=[x_1, \dots, x_n]$ and their corresponding spatial lags (or weighted mean values for the geographical contiguous areas) measured in terms of their deviations from their average. The global univariate Moran's *I* is given by:

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n v_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n v_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad i \neq j \quad \text{or} \quad I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n v_{ij} z_i z_j}, \quad i \neq j$$

where $V=[v_{ij}]$ is the spatial weight matrix, which is a measure of the spatial proximity between municipality *i* and *j*; and $z'_x = [z_i] = [x_i - \bar{x} / \sigma_x]$ a vector of the centred values of *x* that have been normalised (standardised). In significance testing of the null hypothesis of spatial randomness or no spatial autocorrelation one can assume that one is sampling from areas whose spatial process realisations are normally distributed with constant mean and constant variance for each area. Otherwise, a randomisation approach is implemented.

Lee (2001) also proposed a global univariate spatial association measure given by:

$$S(x) = \frac{n}{\sum_{i=1}^n \left(\sum_{j=1}^n v_{ij} \right)^2} \frac{\sum_{i=1}^n \sum_{j=1}^n v_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad i \neq j \quad \text{or} \quad S(x) = \frac{n}{\sum_{i=1}^n \left(\sum_{j=1}^n v_{ij} \right)^2} \frac{\sum_{i=1}^n \sum_{j=1}^n v_{ij} z_i z_j}{\sum_{i=1}^n z_i^2}, \quad i \neq j$$

The only difference between these two measures is that the spatial weights used in Lee's *S* is row-sum standardised,

$$v_{ij} / \sum_{i=1}^n v_{ij}, \text{ while that in Moran's } I \text{ is double standardised } v_{ij} / \sum_{i=1}^n \sum_{j=1}^n v_{ij}.$$

These univariate spatial autocorrelation measures were calculated for the three cardiovascular conditions and compared using inferential properties of the significance testing using Monte-Carlo simulations.

Local indices of spatial autocorrelation

Having established the presence of an underlying pattern or spatial clustering in the data using Global measures such as the Moran's *I* discussed in the section above, one may be interested in detecting hotspots of increased rates or coldspots of reduced rates that could have caused the global statistic to be significant. Furthermore, one can also identify outliers using LISA. Hotspots and coldspots are associated with positive spatial autocorrelation. Outliers are identified when the sign for local spatial autocorrelation negates that of the global spatial autocorrelation. For instance, when the global statistics are saying there is positive spatial autocorrelation then local areas with negative spatial autocorrelations will be spatial outliers. Although there are various LISA applications (Anselin, 1995) here we only consider the most widely used local Moran's *I* statistic given by:

$$I_i = \frac{n(x_i - \bar{x}) \times \sum_{j=1}^n v_{ij} (x_j - \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2}, \quad i \neq j \quad \text{where} \quad \frac{1}{n} \sum_{i=1}^n I_i = I.$$

Limitations of univariate spatial autocorrelation measures

Univariate spatial autocorrelation measures come short whenever there is a need to investigate spatial dependence between two or more interrelated diseases. In univariate spatial autocorrelation analysis it means that one is assuming that the response variable is influenced only by itself in its immediate neighbourhood. By finding the linear relationship between a variable and its spatial lag, then one is looking at testing the hypothesis that as the disease outcome in an area rises then the disease outcome also rises in the nearby areas. The assumption here is that the only factor influencing the rise of a disease outcome in an area is the rise of the disease outcome among its nearest neighbours. But in reality, a rise in disease outcome in an area is due to a multiple of factors. In the case of interrelated diseases like CVDs it may be due to the presence of common biomarkers or related disease outcomes in the nearby areas. Univariate spatial autocorrelations are not able to establish such relationships, thus the need to consider bivariate and multivariate spatial autocorrelation measures.

Bivariate spatial autocorrelation measures

Bivariate spatial association measures were used in this study to test spatial dependence between two diseases as well as to test if there is a difference in the spatial distribution of a disease over two-time points. The bivariate methods applied in this study are variants of the formulation by Wartenberg (1985) and were derived using the popular Moran's *I* univariate spatial autocorrelation measure to detect clustering for one disease. The method suggested by Wartenberg (1985) for extending Moran's *I* to multivariate spatial analysis involves the derivation of a matrix of bivariate spatial autocorrelations. This matrix is, in turn, analysed using spatial principal component analysis (sPCA) resulting in a set of spatial factors that represent the total spatial pattern. While it is preferable to use the row-sum standardised weights in the formulation of Wartenberg (1985), it was found to be problematic as it leads to an asymmetric matrix of bivariate spatial association measures to be diagonalised which is complex to solve, as finding eigenvalues of such a matrix is difficult (Lee, 2001).

Lee (2001) gave conditions that must be satisfied by a bivariate spatial association (SAC) measure to be used for diagonalisation: a bivariate SAC measure must be a function of the respective individual univariate spatial autocorrelations and the *point to point* correlation of some sort between the two variables as measured by Pearson's correlation coefficient. Lee (2001) used the idea of a first order spatial lag (the weighted mean values for the immediate neighbours *j* of an area *i*) given by $Lx_i = \sum_{j=1}^n w_{ij} \cdot x_j = \bar{x}$, to show that the Moran's *I*, when applied with a row-sum standardised matrix, can be rewritten as:

$$I_x = \sqrt{\frac{\sum_{i=1}^n (\bar{x}_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \cdot \sqrt{\frac{\sum_{i=1}^n (\bar{x}_i - \bar{x})^2}{\sum_{i=1}^n (\bar{x}_i - \bar{x})^2}} \cdot r_{x, \bar{x}} \cong \sqrt{SSS_x} \cdot r_{x, \bar{x}}$$

where SSS_x is a spatial smoothing scalar. It implies that the

Moran's I is a product of a spatial smoothing scalar, SSS_x , and the correlation of a variable and its spatial lag, and can be written as $I_x = I_{x,\bar{x}}$.

Deductively, the bivariate Moran's I between two variables X and Y was shown to be:

$$I_{x,y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} \cong \sqrt{SSS_y} \cdot r_{x,\bar{y}} \quad \text{Eq. 1}$$

Eq. 1 is a product of a spatial smoothing scalar (SSS) of a variable and the correlation of the variable and the spatial lag of the other variable. Clearly, the bivariate Moran's I does not satisfy the conditions set out by Lee (2001) as it is a function of only one univariate spatial association measure and a *point-to-point* association of two variables. Thus, Lee (2001) concluded that Wartenberg's (1985) formulations are inadequate and should not be used in multivariate analysis. He went on to derive a bivariate spatial autocorrelation measure for use as a basis for multivariate spatial analysis:

$$L_{x,y} = \frac{n}{\sum_{i=1}^n (\sum_{j=1}^n w_{ij})^2} \cdot \frac{\sum_{i=1}^n \left(\sum_{j=1}^n w_{ij} (x_i - \bar{x}) \right) \left(\sum_{j=1}^n w_{ij} (y_i - \bar{y}) \right)}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} \cong \sqrt{SSS_y} \cdot \sqrt{SSS_x} \cdot r_{\bar{x},\bar{y}} \quad \text{Eq. 2}$$

Eq. 2 by Lee (2001), known as Lee's L , is not only in line with his conditions for a bivariate SAC measure, but also produces a symmetric bivariate SAC matrix to be used for deriving total multivariate spatial autocorrelations. The Pearson's correlation part of Lee's derivation, $r_{\bar{x},\bar{y}}$, is between the spatial lags of the two variables that will be considered. One problem that may arise is that this correlation between the lagged values may differ significantly from that between the original values of the two variables and may have different signs (Dray *et al.*, 2008; Lee, 2001). This was not the case with the data used in this study as correlations between lagged values did not have different signs and had close values to the original values.

Additionally, Lee (2001) showed that if $X=Y$ then:

$$L_{x,x} = \sqrt{SSS_x} \cdot \sqrt{SSS_x} \cdot \underbrace{r_{\bar{x},\bar{x}}}_{=1} = SSS_x = S_x \quad \text{Eq. 3}$$

Eq. 3 is often referred to as Lee's S and can be used to measure univariate spatial autocorrelation just like Univariate Moran's I (Lee, 2004).

In spite of the criticism of the ideas of Wartenberg (1985), the approach has remained popular, with Anselin *et al.* (2002) expanding the formulation to visual analysis of bivariate Moran's I spatial association measure. This expansion was done for both global and local indexes using a standardised weight matrix (W). This bivariate measure does not meet the conditions set out by Lee (2001) as noted earlier. In order to overcome this difficulty, Dray *et al.* (2008) noted that instead of using W in his formulations, Lee (2001) should have used $W+W^T$, as originally suggested by De Jong *et al.* (1984). Dray *et al.* (2008) then went on to use the transformation by De Jong *et al.* (1984) to develop a bivariate spatial association measure:

$$H_{x,y} = \frac{1}{2} \left[\sqrt{SSS_x} \cdot r_{\bar{x},y} + \sqrt{SSS_y} \cdot r_{y,\bar{x}} \right] \quad \text{Eq. 4}$$

The bivariate spatial association measure in Eq. 4 is not only symmetric, but satisfies the conditions of Lee (2001). In addition, the measure is a function of the correlation of one variable and the lag of the second variable, thus indirectly connecting it to the regression formulations by Anselin *et al.* (2002). The global univariate Moran's I and Lee's S were calculated to determine the presence of clustering for each of the three CVDs of HHD, CVA and IHD in South Africa, while global bivariate association was determined using bivariate Moran's I , Lee's L and Dray's H . A comparison was made to establish if the three bivariate association measures give similar results based on the significance test. The LISA were used to determine the clusters and co-clusters which were, in turn, mapped on choropleth maps. A full description of LISA methods has been given elsewhere (Anselin *et al.*, 2002) and is not included in this paper. A permutation approach was used to determine the significance of the test, and the framework of this significance testing is discussed in the next section.

Framework for significance testing

A total of 99 permutations were run during significance testing for each of the test statistics to determine the presence of global spatial autocorrelations. Research has shown that 99 permutations should suffice when testing at 5% significance level. Lee (2004) provided a framework for significance testing of indicators of spatial association measures. In this framework an indicator of spatial association measure, say $\Gamma(obs)$, was determined for the deaths due to the three cardiovascular conditions. Simulation was done 99 times, by randomly assigning observed values to the locations and each time determining Γ . The 99 Γ values were then arranged in ascending order and the number of times, say l , that $\Gamma > \Gamma(obs)$ was determined. The P-value was then determined by finding $P(\Gamma > \Gamma(obs))$. This is equivalent to dividing l by 100, where the one that was added to the 99 simulations to get 100 is for the observed statistic, $\Gamma(obs)$. The results are provided in the next section.

Results

The quantile maps of crude and smoothed mortality rates at municipal level, for each of the three studied CVD conditions in South Africa for the years 2001, 2007 and 2011 are shown in Figures 1 and 2. Municipalities in the upper quantile indicate areas that experienced high rates of mortality while those in the lower quantile indicate areas with low rates of mortality. The higher the quantile the darker the colour (ranging from quantile 1 in white to quantile 5 in black). It follows that areas with the darker shade indicate areas of higher mortality than a relatively lighter shade.

Generally, the maps of variations in CVD condition show some clear clustering; for example, areas with higher risk of CVA and IHD deaths are in the centre of the country stretching to the south west part of the country while areas with higher risk of HHD mortality are numerous pockets of clusters in the central part of the country.

Figure 2 shows the quantile maps of the same CVDs in Figure 1 but now using EB rates. The effects of smoothing are clear with CVA and IHD showing even darker shade from the middle of the country stretching to South West of the country. The HHD01 and

HHD11 EB rates maps are also showing a more continuous, darker shade in the middle of the country stretching to the south of the country than the raw rates quantile maps revealing potential clusters of high hypertension risk in these regions. It seems the effects of stabilising the crude rates with the EB approach has been, based on the evidence of the quantile maps, an improved ability to discern areas of higher risk. A fact that was confirmed in the next section by statistical significance tests of spatial autocorrelations. Unfortunately, the SMRs maps (not shown here) did not do the same especially with CVA maps becoming lesser dark in areas of potential clustering.

Univariate cluster analysis

Univariate global spatial autocorrelation

The previous assessment of geographical variations for both crude and smoothed rates has shown evidence of clustering in CVD outcomes. In order to formally investigate spatial association, we measured the association in a formal way by using univariate (in this section) and bivariate clustering statistics (see *Bivariate analysis* section for details). Table 2 presents the derived values for each CVD for the whole of South Africa for the years 2001, 2007 and 2011. For comparison purposes, the derivations

were done using raw rates and EB rates and indirectly standardised rates. CVA_{zy} , IHD_{zy} and HHD_{zy} represent mortality due to cerebrovascular, ischaemic and hypertensive heart conditions in the year zy , respectively. Here zy takes values 01, 07 and 11, representing the years 2001, 2007 and 2011, respectively.

The univariate Moran's I and Lee's S in Table 2 both confirm that the distribution of the three CVDs of IHD, CVA and IHD vary geographically, using the three rates. This geographic variation is highly significant for EB rates with P-values less than 0.001 and is persistent over the ten-year period for all three conditions. In all the cases, the calculated statistics for both Lee's S and Moran's I are all positive and significant across the years except for CVA01 SMRs map which is insignificant. This means that the likelihood of the spatial patterns generated by mortality due to each of the three CVDs being due to random chance is negligibly small (less than 5%). Thus, one can conclude that the probability is high that municipalities that are nearer to each other have a tendency of having comparable baseline mortality rates than the distant municipalities. In other words, there is some form of clustering exhibited by all three CVDs at the 5% significance level. The EB rates have the highest Lee's S and Moran's I values suggesting more clustering for the EB data than the raw data. This is a reflection of what is seen in the quantile maps in Figures 1 and 2.

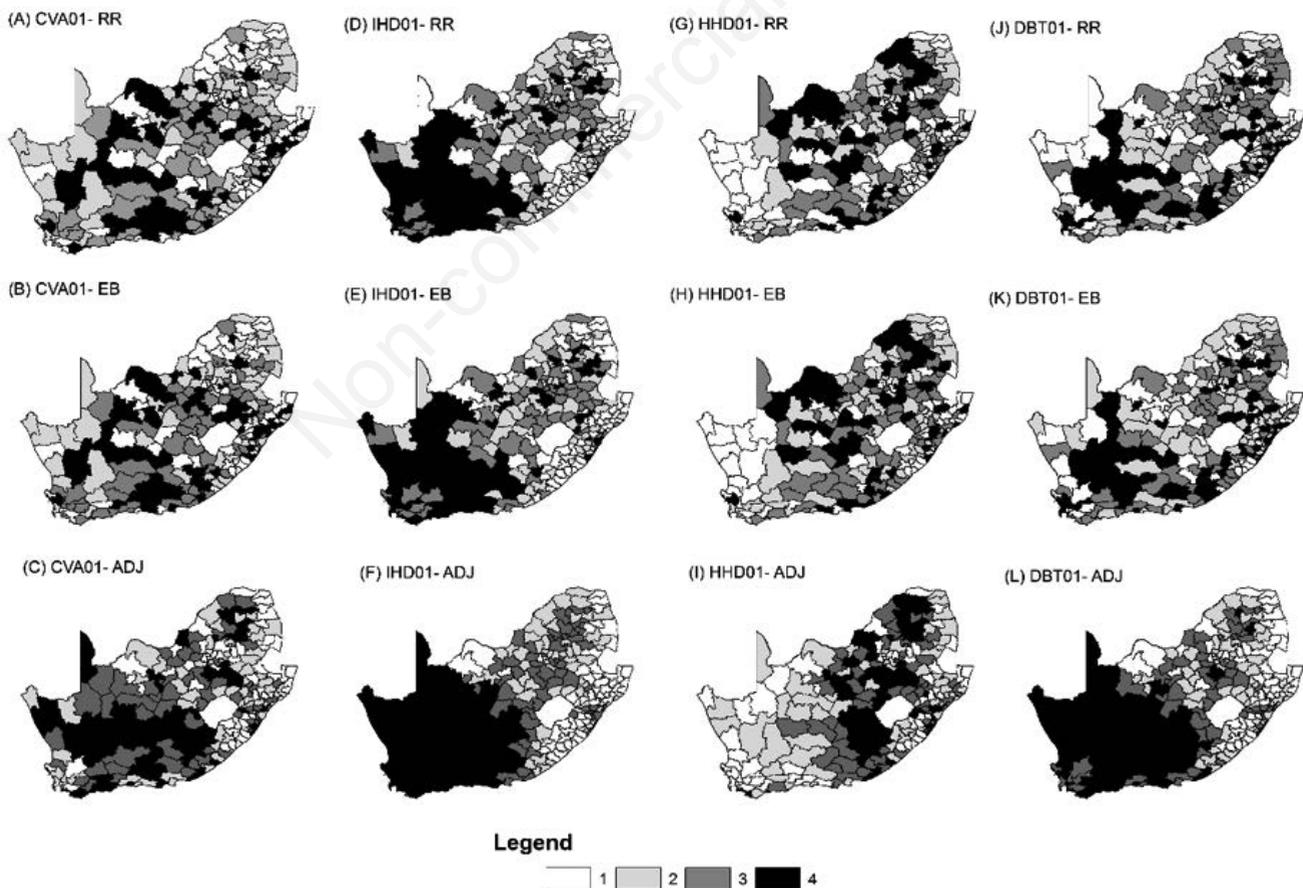


Figure 1. Quartile maps showing the distribution of raw mortality rates due to each of the three cardiovascular conditions for the years 2001, 2007 and 2011.

Univariate hotspot analysis

The local indicators of autocorrelation (LISA) based on Moran's I were used to determine the actual clusters at municipal level. The resulting univariate LISA maps for raw rates and EB rates are shown in Figures 3 and 4, respectively. Hotspots, which are municipalities of high mortality incidences that are surrounded by municipalities with high mortality incidences, are indicated by

a *High-High* key on the map, while the coldspots, which are municipalities of low mortality incidences that are surrounded by municipalities with low mortality incidences, are indicated by a *Low-Low* key. In addition, there are outliers indicated by *High-Low*, which are municipalities of high mortality incidences that are surrounded by municipalities with low mortality incidences, and *Low-High*, which are municipalities of low mortality incidences

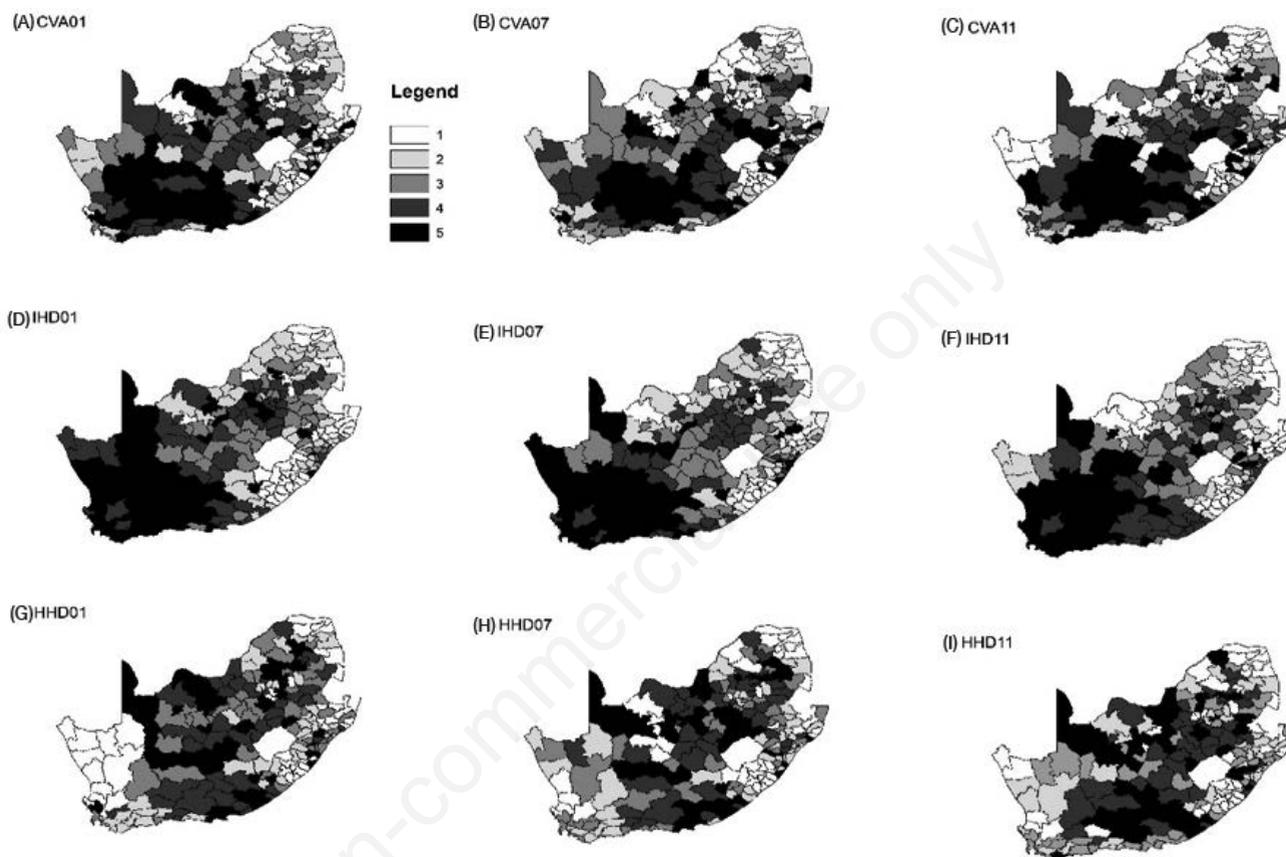


Figure 2. Quartile maps showing the distribution of empirically Bayes smoothed mortality rates due to each of the three cardiovascular conditions for the years 2001, 2007 and 2011.

Table 2. Univariate global spatial autocorrelations for the mortality rates due to cerebrovascular heart disease, ischaemic heart disease and hypertensive heart disease in 2001, 2007 and 2011.

Variable	2001			2007			2011		
	Raw rate	Lee's S EB rate	SMR	Raw rate	Lee's S EB rate	SMR	Raw rate	Lee's S EB rate	SMR
CVA01	0.298***	0.345***	0.264***	0.116**	0.189***	0.058INS	0.125**	0.157**	0.092**
CVA07	0.289***	0.327***	0.326***	0.100**	0.191***	0.155**	0.408***	0.466***	0.295***
CVA11	0.313***	0.332***	0.288***	0.411***	0.531***	0.295***	0.351***	0.426***	0.239***
IHD01	0.576***	0.627***	0.490***	0.408***	0.466***	0.295***	0.351***	0.426***	0.239***
IHD07	0.529***	0.605***	0.437***	0.411***	0.531***	0.295***	0.351***	0.426***	0.239***
IHD11	0.478***	0.528***	0.418***	0.351***	0.426***	0.239***	0.147**	0.300***	0.170***
HHD01	0.316***	0.401***	0.338***	0.147**	0.300***	0.170***	0.331***	0.410***	0.319***
HHD07	0.331***	0.410***	0.319***	0.169**	0.291***	0.152**	0.317***	0.373***	0.318***
HHD11	0.317***	0.373***	0.318***	0.163**	0.253***	0.169**			

CVA, cerebrovascular heart disease; IHD, ischaemic heart disease; HHD, hypertensive heart disease; EB, empirically Bayes; SMR, standardised mortality rates. *** $P < 0.001$; ** $P < 0.05$; INS, no significant association at 5% level.

that are surrounded by municipalities with high mortality incidences. The hotspots are of major concern as they represent clusters of high risk of mortality due to the CVDs.

It can be seen in Figure 4A-C that the number of municipalities with CVA hotspots have slightly reduced with hotspots being found more towards the middle of the country and less easterly. The cluster of high CVA mortality hotspots span over four provinces: Northern Cape, KwaZulu-Natal, Free State and Eastern Cape. The district with the most identified hotspots of high CVA mortality in 2011 is Pixley ka Seme in Northern Cape with six identified hotspots (Emthanjeni, Kareeberg, Siyathemba, Thembelihle, Umsobomvu, Ubuntu).

Figure 4D-F shows the change in IHD clusters over time. These LISA maps reveal that the spatial dynamics of IHD are stable over the period under study, with hotspots located in the centre and spanning all the way to the south-west coast of the country. Districts with municipalities hit hard by IHD deaths are Central Karoo, Cape Winelands, Eden, Overberg, West Coast (Western Cape Province), Cacadu (Eastern Cape), Namakwa and Pixley ka Seme (Northern Cape).

The LISA analysis for HHD, shown in Figure 4G-I, depicts unstable spatial dynamics over time with respect to hotspots but coldspot clusters have remained fairly constant along the eastern coast and north-eastern part of the country. There is no clear expla-

nation for this phenomenon at this point in time. To aid the analysis of Figures 3 and 4, the number of municipalities forming clusters (hotspots or coldspots) are presented in Table 2. Only hotspots and coldspots are shown as the other clusters (high-low and low-high) are very few and do not change the general trend.

There are more municipalities forming clusters when EB rates are used than when using raw rates. This explains why Lee's S and Moran's I values for raw rates are smaller than those derived from EB rates. This reflects, once again, the effect of EB smoothing on increased detection of mortality risk in an area. The stabilising effect of EB smoothing is more evident in the number of hotspots formed for each year for HHD. When raw rates are used the number of hotspots jumps from 8 in 2001 to 16 in 2007 (a 100% jump) then drops again to 8 in 2011 (a 50% drop). But when EB rates are used the number of HHD hotspots in the country are almost constant over time moving up by 11% from 18 in 2001 to 20 in 2007 before decreasing by 20% to 16 in 2011.

Bivariate analysis

Mapping bivariate association of individual cardiovascular diseases over time

Table 2 shows that geographical variation of mortality due to each CVD is highly significant over the years and the clusters have

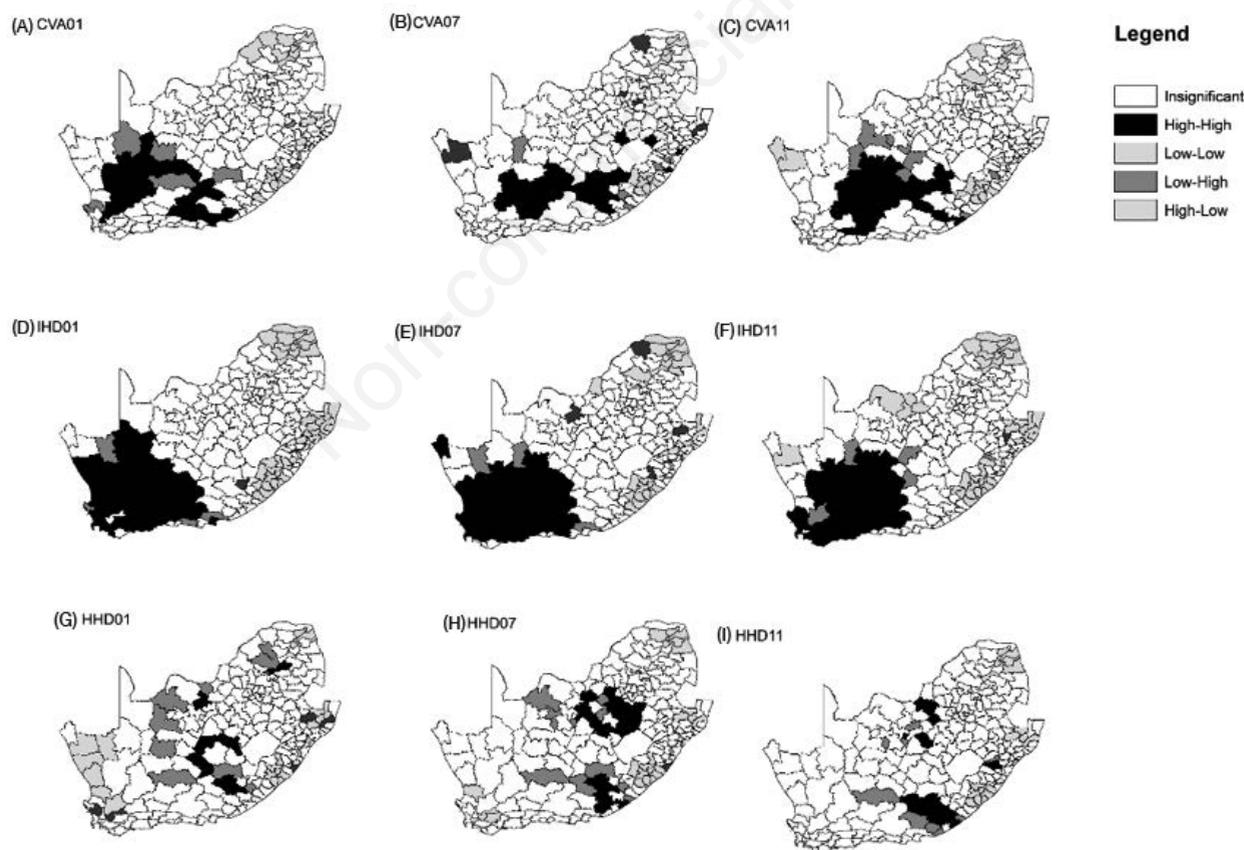


Figure 3. Trends in raw rates based univariate Moran's I local indices of spatial autocorrelation clusters for each of the cardiovascular conditions for the years 2001, 2007 and 2011.

been shown in Figure 4 using EB data. It has not been shown, however, whether or not the variation in the distribution of each CVD risk is the same over the years. In this section, bivariate spatial autocorrelation measures are used to determine if there is a difference in spatial distribution of mortality risk due to each of the CVDs between two time points. This will shed light on whether or not the geographical distribution of individual CVDs is changing over time. The stability of spatial dynamics in the distribution of a disease is important if there is to be some semblance of predictability. This may be helpful when deciding the course of action to be taken when faced with an epidemic as intervention programmes are conceived. Changing spatial dynamics in the distribution of the disease may make it quite complex to contain the disease. Generally, one would not expect spatial dynamics of CVDs to change much within a short period of time as the factors and habits contributing to the emergence of these diseases take time to control. Thus, one would expect the bivariate spatial dependency of the distribution of each CVD over two time periods to be significantly positive.

The dependence of CVD rates in space for each of the studied conditions was tested for two different periods using bivariate Lee's L , Moran's I and Dray's H . The analysis was conducted on each CVD for the following comparative periods: 2001 vs 2007; 2001 vs 2011; and 2007 vs 2011. Results are provided in Table 3.

Generally, the three indicators of bivariate spatial autocorrelation show that there is significant spatial dependency in how each

Table 3. Number of municipalities forming hotspots and coldspots clusters for the mortality rates due to cerebrovascular heart disease, ischaemic heart disease and hypertensive heart disease in 2001, 2007 and 2011.

Variable	Rate type	Hotspots	Coldspots	Total
CVA01	Raw Rate	12	18	30
	EB Rate	16	23	39
CVA07	Raw Rate	14	9	23
	EB Rate	14	22	36
CVA11	Raw Rate	15	13	28
	EB Rate	14	17	31
IHD01	Raw Rate	29	39	68
	EB Rate	36	41	77
IHD07	Raw Rate	30	28	58
	EB Rate	38	33	71
IHD11	Raw Rate	27	41	68
	EB Rate	30	46	76
HHD01	Raw Rate	8	21	29
	EB Rate	18	26	44
HHD07	Raw Rate	16	23	39
	EB Rate	20	27	47
HHD11	Raw Rate	8	18	26
	EB Rate	16	21	37

CVA, cerebrovascular heart disease; IHD, ischaemic heart disease; HHD, hypertensive heart disease; EB, empirically Bayes.

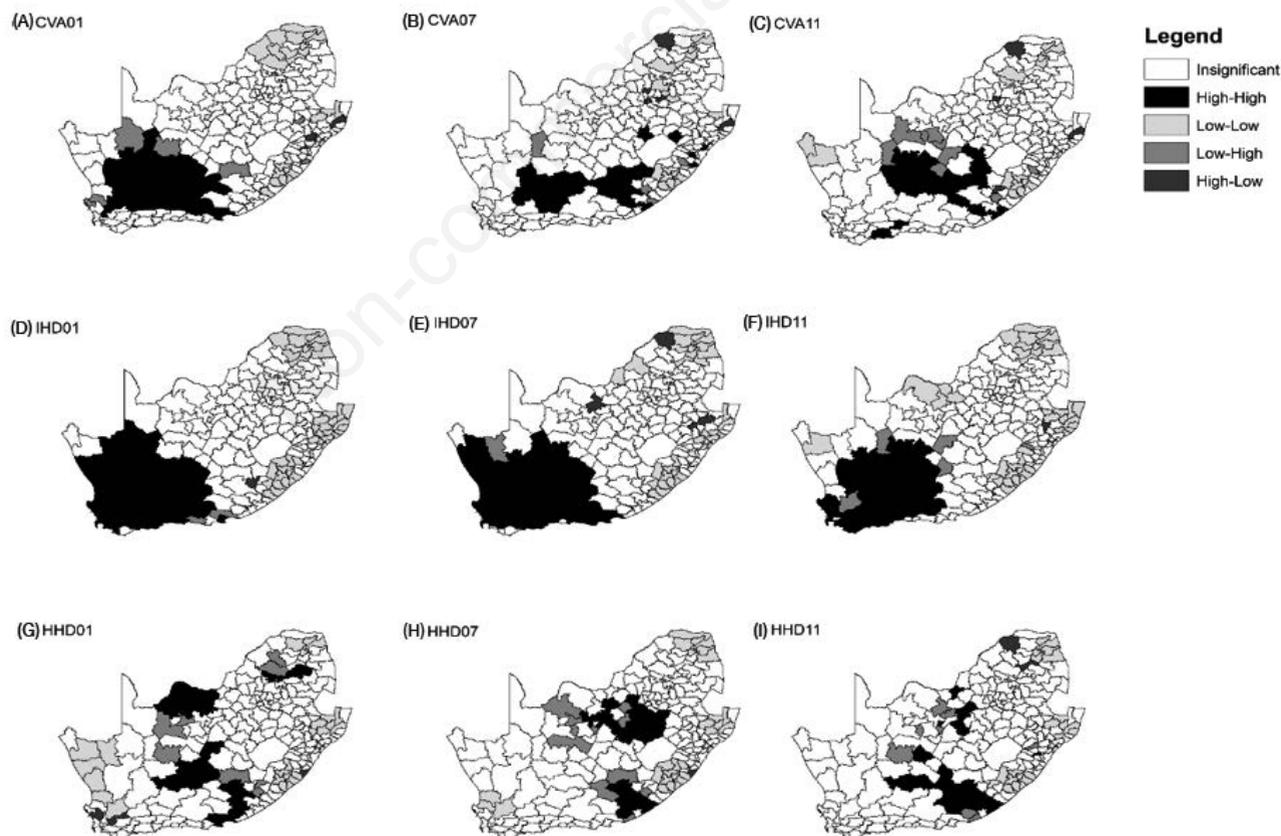


Figure 4. Trends in empirically Bayes rates based univariate Moran's I local indices of spatial autocorrelation clusters for each of the cardiovascular conditions for the years 2001, 2007 and 2011.

disease is spatially distributed between two time periods. It can, thus, be concluded that the spatial distribution of the risk of mortality due to each CVD has not significantly changed over the course of the ten-year period under review. The EB rates show higher values of the three bivariate measures than both the raw rate and SMR data. This indicates that bivariate association is more pronounced in maps derived from EB rate data than the other two datasets. The bivariate LISA analyses of the combinations in Table 3 derived from the EB rates are presented in Figure 5.

In Figure 5D, as an example, observed hotspots are areas of high mortality of IHD in 2001 whose neighbourhood in 2007 also exhibit high mortality of IHD to form a co-cluster of high mortality for the two-time points in the south-western part of the country. Figure 5E and F reveals that the joint cluster is stable for the years 2001/2011 and 2007/2011. As for HHD, Figure 5G-I shows stable clusters of coldspots along the east coast and the north-eastern part of the country and some hotspots scattered in the middle of the country.

Bivariate spatial association between two cardiovascular diseases at one point in time

We also looked at determining spatial dependency between

two different CVDs at a cross-section. One can hypothesise that CVDs should co-cluster or show spatial dependency at a point in time as they share risk factors. Table 4 presents the bivariate association measure values calculated for the possible combinations of the three CVDs for the years 2001, 2007 and 2011 to determine spatial dependence based on raw, indirectly standardised and EB rates data.

The bivariate Moran's I and Dray's H show similar results. This is not surprising as the methods are based on the same derivation. All three methods agree, based on the EB rates, that there is spatial dependence between the following maps: CVA and IHD for the years 2001, 2007 and 2011; IHD and HHD for the year 2011; and CVA and HHD for the years 2007 and 2011. Lee's bivariate L differs with the other two methods when EB rates are used only once: CVA01-HHD01. In this instance Lee's L reveals highly significant spatial dependency between the two maps while the Moran's I says it is insignificant. There are more of such instances when raw rates and SMRs are used (see CVA-HHD). Here the raw rates and indirectly SMRs fail to reveal spatial dependency between CVA and HHD maps for the years 2001, 2007 and 2011 when Moran's I is used but says otherwise when Lee's L is used. Furthermore, Lee's L obtained similar results for raw, SMR and EB

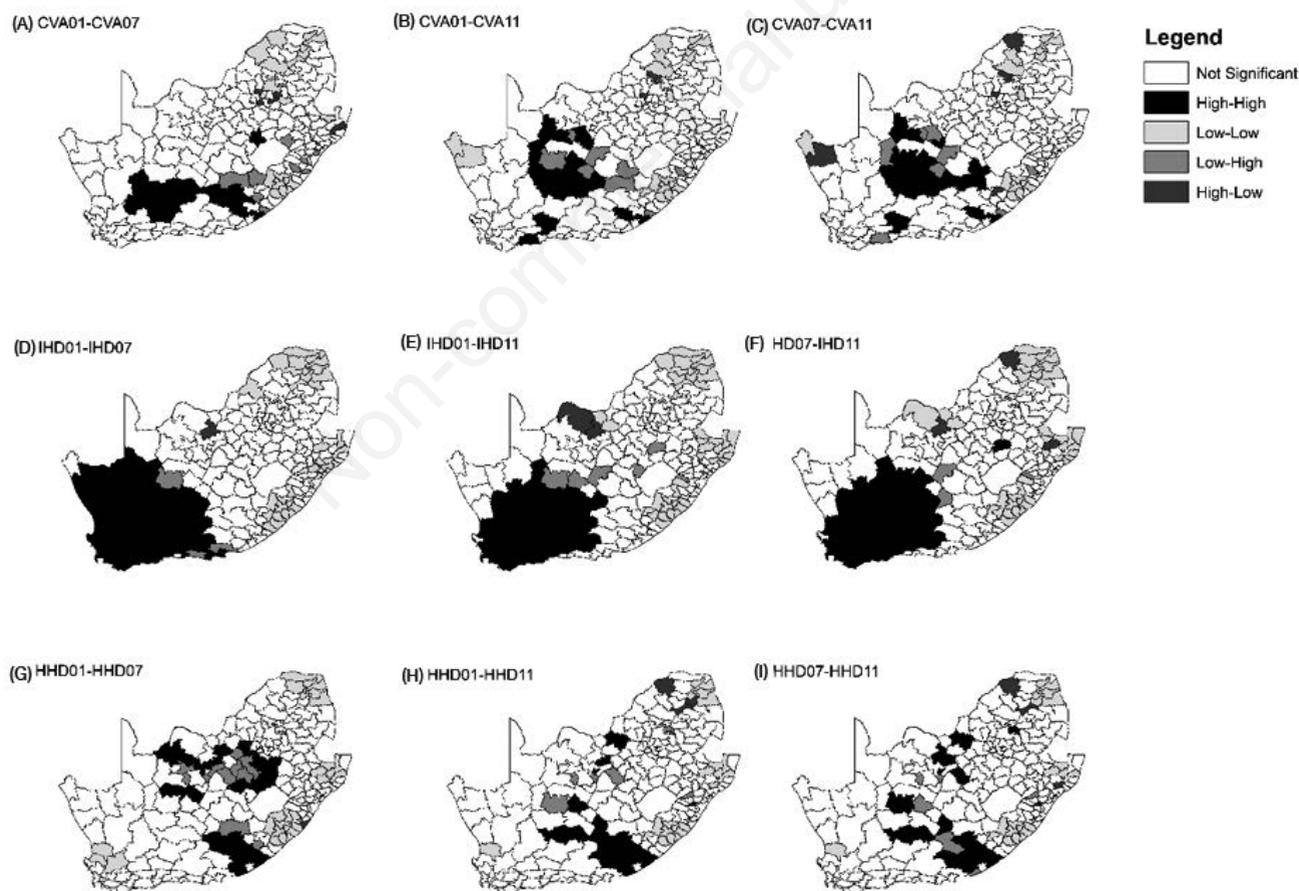


Figure 5. The empirically Bayes rates based Bivariate Moran's I clusters between cerebrovascular heart disease, ischaemic heart disease and hypertensive heart disease deaths for the years 2001, 2007 and 2011.

smoothed data except for the relationship between CVA07 and IHD07 for which the SMR is insignificant. All this evidence suggest that the Lee's L is more robust to small changes in the spatial data being. It also seems Lee's L is more adept at detecting co-clustering where Moran's I would say otherwise. The two methods

may be used to complement each other such as when we chose to use EB rates where both methods are generally in agreement.

The joint local hotspots of the CVDs, based on EB data, are shown in Figure 6. The CVA-IHD maps show the most pronounced clusters of hotspots and coldspots. The hotspots (municipalities of

Table 4. Bivariate global spatial autocorrelations measuring spatial dependence of individual cardiovascular disease rates between two time periods for the years 2001, 2007 and 2011.

Association(X-Y)		Lee 2001 ($L_{X,Y}$)			Anselin 2002 ($I_{X,Y}$)			Dray 2008 ($H_{X,Y}$)		
X	Y	Raw rate	EB rate	SMR	Raw rate	EB rate	SMR	Raw rate	EB rate	SMR
CVA01	CVA07	0.171***	0.201***	0.147***	0.063**	0.116**	0.025INS	0.061 ^{INS}	0.112**	0.029 ^{INS}
CVA01	CVA11	0.163***	0.184***	0.114***	0.125**	0.162**	0.078**	0.119***	0.155***	0.072**
CVA07	CVA11	0.197***	0.221***	0.222***	0.128**	0.170**	0.148**	0.125***	0.168***	0.144***
IHD01	IHD07	0.502***	0.555***	0.407***	0.406***	0.491***	0.293***	0.409***	0.492***	0.306***
IHD01	IHD11	0.419***	0.458***	0.323***	0.388***	0.437***	0.268***	0.383***	0.445***	0.275***
IHD07	IHD11	0.405***	0.473***	0.321***	0.377***	0.458***	0.276***	0.374***	0.459***	0.272***
HHD01	HHD07	0.231***	0.315***	0.211***	0.160**	0.271***	0.143**	0.160***	0.268***	0.141***
HHD01	HHD11	0.198***	0.276***	0.167***	0.154**	0.269***	0.149**	0.146***	0.258***	0.134**
HHD07	HHD11	0.258***	0.321***	0.245***	0.200***	0.291***	0.197***	0.191***	0.284***	0.188***

CVA, cerebrovascular heart disease; IHD, ischaemic heart disease; HHD, hypertensive heart disease; EB, empirically Bayes; SMR, standardised mortality rates. *** $P < 0.001$; ** $P < 0.05$; ^{INS}, no significant association at 5% level.

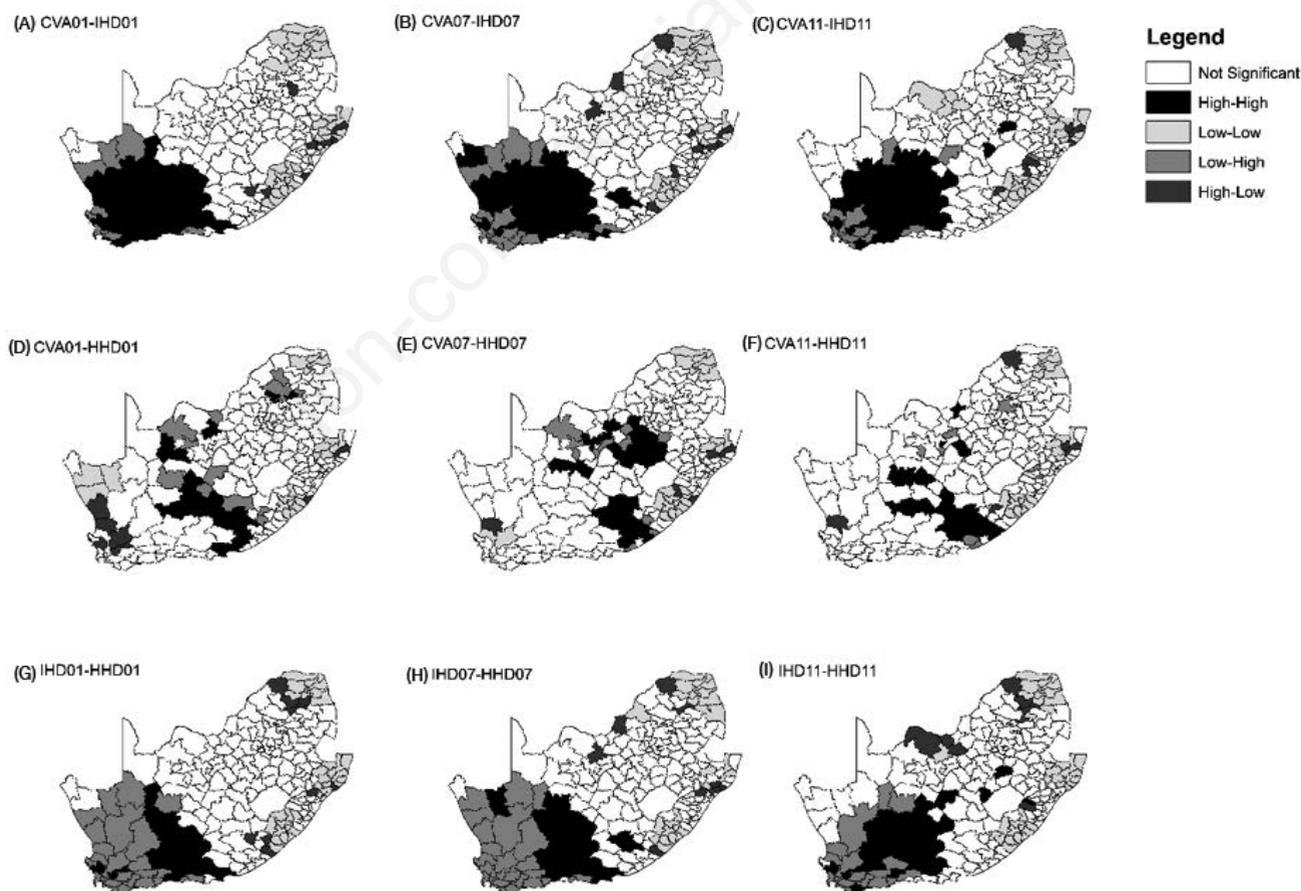


Figure 6. The empirically Bayes rates based Bivariate Moran's I local indices of spatial autocorrelation maps between two cardiovascular disease at a point in time for the years 2001, 2007 and 2011.



high CVA deaths that are surrounded by municipalities of high IHD deaths) are located at the centre of the country spreading south-west of the country. The co-clustering is generally concentrated in the sub-districts of the Western Cape Province, Northern Cape and Eastern Cape Province with Eden, Pixley ka Seme and Central Karoo having the highest number of municipal joint hotspots for the year 2011. Co-clustering of high mortality between HHD and IHD occurs in the centre of the south-western part of the country, while coldspots are on the eastern coast and north-east of the country.

Discussion

This study has analysed very important health data concerning mortality in South Africa to assess spatial dependence in mortality outcomes attributed to the emerging and rising CVDs. The study has used novel spatial statistical techniques to measure both global and local spatial clustering of multiple CVD outcomes jointly across municipalities. This has offered better insights into the disease aetiology over the usual univariate spatial clustering measures.

There was a high level of heterogeneity exhibited in the spatial distribution of mortality rates due to each of the three CVDs studied. In these cases, on one hand, the inequalities in the risk of dying gave rise to hotspot municipalities with high levels of incidence of mortality surrounded by municipalities of high incidence of mortality, and on the other hand, they gave rise to coldspots of low incidences surrounded by municipalities of low incidences. The LISA maps have shown that CVD hotspots are generally located in the south-west of the country. It is not surprising that CVD deaths are located in these areas as reports by Stats SA have previously shown that IHD was the leading cause of deaths in Namakwa District of the Northern Cape in 2007 (Statistics South Africa, 2009) while CVA was reported as the leading cause of deaths in Eden and Overberg Districts in 2011 (Statistics South Africa, 2014). In addition, Groenewald *et al.* (2014) showed that the NCD problematic districts with YLLs between 45% and 56% are all located in the south-west of the country, and these are Namakwa,

West Coast, Cape Winelands, Overberg and the City of Cape Town. Those that had YLLs between 39% and 45% are in different provinces and include, Central Karoo and Pixley ka Seme, also in the south-west of the country. It is in these districts and surrounding districts that the hotspots of CVDs and their co-clusters have been detected by this study. Additionally, the work by Kandala *et al.* (Kandala *et al.*, 2014) revealed that cardiovascular diseases are distributed in the south-west part of the country, which is in line with our findings.

Bivariate analysis was used to answer the two main objective functions of whether two interrelated diseases co-cluster and to test if the spatial patterns of a disease differ between two time points. The results of bivariate analysis revealed that the spatial distribution of mortality risk attributed to CVA and IHD has been stable over the years with minimal changes. Thus, the government and other civil organisations may want to formulate intervention programmes knowing fully well that the identified clusters of high risk will still be there unless some intervention of some sort has been done. In addition, bivariate analysis has revealed significant co-clustering between CVA and each of the other two CVD conditions (IHD and HHD), but co-clustering between IHD and HHD was found to be significant for the year 2011 only. Thus, any integrated approach that can be mooted will need to target the high mortality risk joint clusters of either CVA and IHD or CVA and HHD or IHD and HHD or all of them simultaneously. These joint clusters, presented in Figure 6, are found in the central part of the country for CVA and HHD and in the central part of the country stretching to the south-west part of the country for CVA and IHD and also for IHD and HHD (2011).

The bivariate Moran's I and Dray's H are direct extensions of the approach by Wartenberg (1985), and they both generally gave similar results. Thus, one can use Dray's bivariate association measure in bivariate analysis in the knowledge that it can yield similar results to the original approach by Wartenberg (1985) while at the same time addressing the flaws of the latter as espoused by Lee (2001). Results in this study suggest using Lee's approach and Moran's approach to complement each other. The next step in this study is to develop multivariate spatial clustering measures that can be used for three or more health outcomes.

Table 5. Bivariate global spatial autocorrelations between two cardiovascular diseases at a point in time for the years 2001, 2007 and 2011.

Association(X-Y)		Lee 2001 ($L_{X,Y}$)			Anselin 2002 ($I_{X,Y}$)			Dray 2008 ($H_{X,Y}$)		
X	Y	Raw rate	EB rate	SMR	Raw rate	EB rate	SMR	Raw rate	EB rate	SMR
CVA01	IHD01	0.296***	0.338***	0.188***	0.184***	0.228***	0.074**	0.194**	0.241***	0.078**
CVA07	IHD07	0.144**	0.171***	0.040 ^{INS}	0.056 ^{INS}	0.103**	-0.055 ^{INS}	0.068 ^{INS}	0.117**	-0.046 ^{INS}
CVA11	IHD11	0.255***	0.261***	0.142***	0.142**	0.166**	0.007 ^{INS}	0.144**	0.166**	0.009 ^{INS}
CVA01	HHD01	0.130***	0.148***	0.135***	0.011 ^{INS}	0.057 ^{INS}	0.002 ^{INS}	0.012 ^{INS}	0.055 ^{INS}	0.004 ^{INS}
CVA07	HHD07	0.127***	0.176***	0.123***	0.049 ^{INS}	0.116**	0.050 ^{INS}	0.052 ^{INS}	0.122**	0.050 ^{INS}
CVA11	HHD11	0.165***	0.200***	0.140***	0.053 ^{INS}	0.107**	0.030 ^{INS}	0.050 ^{INS}	0.106**	0.028 ^{INS}
HHD01	IHD01	0.056 ^{INS}	0.059 ^{INS}	0.024 ^{INS}	-0.026 ^{INS}	-0.008 ^{INS}	-0.060 ^{INS}	-0.036 ^{INS}	-0.013 ^{INS}	-0.066 ^{INS}
HHD07	IHD07	0.032 ^{INS}	0.075 ^{INS}	-0.005 ^{INS}	0.007 ^{INS}	0.045 ^{INS}	-0.042 ^{INS}	0.011 ^{INS}	0.050 ^{INS}	-0.036 ^{INS}
HHD11	IHD11	0.108***	0.130**	0.072**	0.064**	0.097**	0.012 ^{INS}	0.061 ^{INS}	0.096**	0.012 ^{INS}

CVA, cerebrovascular heart disease; IHD, ischaemic heart disease; HHD, hypertensive heart disease; EB, empirically Bayes; SMR, standardised mortality rates. *** $P < 0.001$; ** $P < 0.05$; ^{INS}, not significant association at 5% level.

Conclusions

The results of this study have shown that it is possible to identify clusters and co-clusters of high risk CVD mortality. Thus, using both univariate and bivariate spatial autocorrelation measures in identifying high risk clusters of CVD mortality is a viable strategy for monitoring, prevention and control of CVDs individually and jointly in South Africa. An integrated approach in the diagnosis, treatment and care of patients suffering from diverse cardiovascular conditions may reduce the effects of a number of NCD related mortality in the municipalities that are at elevated risk of mortality attributed to the three diseases.

References

- Alberti KGMM, Zimmet P, 2005. The metabolic syndrome – a worldwide definition. *Lancet* 366:1059-62.
- Anselin L, Syabri I, Smirnov O, 2002. Visualizing multivariate spatial correlation with dynamically linked windows. In: Anselin L, Rey S, eds. *New tools for spatial data analysis: proceedings of the specialist meeting*. Center for Spatially Integrated Social Science, University of California, Santa Barbara, CA.
- Birnbaum JK, Murray CJ, Lozano R, 2011. Exposing misclassified HIV/AIDS deaths in South Africa. *Bull World Health Organ* 89:278-85.
- Bradshaw D, Bourne D, Schneider M, Sayed R, 1995. Mortality patterns of chronic diseases of lifestyle in South Africa. In: Fourie J, Steyn K, eds. *Review of research and identification of essential health research priorities*. South Africa Medical Research Council, Cape Town, South Africa.
- Bradshaw D, Schneider M, Norman R, Bourne D, 2006. Mortality patterns of chronic diseases of lifestyle in South Africa. In: Steyn K, Fourie J, Temple N, eds. *Chronic diseases of lifestyle in South Africa: 1995-2005*. South Africa Medical Research Council, Burden of Disease Research Unit, Cape Town, South Africa.
- Clayton D, Kaldor J, 1987. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* 43:671-81.
- De Jong P, Sprenger C, Van Veen F, 1984. On extreme values of Moran's I and Geary's c . *Geogr Anal* 16:17-24.
- Dray S, Saïd S, Débias F, 2008. Spatial ordination of vegetation data using a generalization of Wartenberg's multivariate spatial correlation. *Ecol Model J Veg Sci* 19:45-56.
- Ford MM, Highfield LD, 2016. Exploring the spatial association between social deprivation and cardiovascular disease mortality at the neighbourhood level. *PLoS One* 11:1-17.
- Groenewald P, Bradshaw D, Day C, Laubscher R, 2014. Burden of disease. In: Massyn N, Day C, Peer N, Padarath A, Barron P, English R, eds. *District Health Barometer: 2013-14*. Health Systems Trust, Durban, South Africa.
- Joubert J, Rao C, Bradshaw D, Vos T, Lopez AD, 2013. Evaluating the quality of national mortality statistics from civil registration in South Africa, 1997-2007. *PLoS One* 5:e64592.
- Kandala NB, Manda SOM, Tigbe WW, Mwambi H, Stranges S, 2014. Geographic distribution of cardiovascular comorbidities in South Africa: a national cross-sectional analysis. *J Appl Stat* 41:1203-16.
- Lee SI, 2001. Developing a bivariate spatial association measure: an integration of Pearson's r and Moran's I . *J Geogr Syst* 4:369-85.
- Lee SI, 2004. A generalized significance testing method for global measures of spatial association: an extension of the Mantel test. *Environ Plann* 36:1687-703.
- Leyland AH, Davies CA, 2005. Empirical Bayes methods for disease mapping. *Stat Method Med Res* 14:17-34.
- Manda SO, Lombard CJ, Mosala T, 2012. Divergent spatial patterns in the prevalence of the human immunodeficiency virus (HIV) and syphilis in South African pregnant women. *Geospat Health* 6:221-31.
- Marshall RJ, 1991. Mapping disease and mortality rates using empirical Bayes estimators. *J R Stat Soc Ser C Appl Stat* 40:283-94.
- Pillay-Van Wyk V, Laubscher R, Msemburi W, Dorrington RE, Groenewald P, Vos T, Matzopoulos R, Prinsloo M, Nojilana B, Nannan N, Somdyala N, Sithole N, Neethling I, Nicol E, Rossouw A, Joubert J, Bradshaw D, 2014. Second South African National Burden of Disease Study: data cleaning, validation and SA NBD List. Burden of Disease Research Unit, South African Medical Research Council, Cape Town, South Africa.
- Sartorius B, Kahn K, Collinson MA, Vounatsou P, Tollman SM, 2010. Space and time clustering of mortality in rural South Africa (Agincourt HDSS), 1992-2007. *Global Health Action* 1:50-8.
- Sartorius BK, Sartorius K, Chirwa TF, Fonn S, 2011. Infant mortality in South Africa - distribution, associations and policy implications, 2007: an ecological spatial analysis. *Int J Health Geogr* 10:1-13.
- Statistics South Africa, 2009. Mortality and causes of death in South Africa, 2007: findings from death notification. Statistical Release P0309.3. Statistics South Africa, Pretoria, South Africa.
- Statistics South Africa, 2014. Mortality and causes of death in South Africa, 2011: findings from death notification. Statistical Release P0309.3. Statistics South Africa, Pretoria, South Africa.
- Tanser F, Barnighausen T, Cooke GS, L, Newell ML, 2009. Localized spatial clustering of HIV infections in a widely disseminated rural South African epidemic. *Int J Epidemiol* 38:1008-16.
- Waller LA, Gotway CA, 2004. *Applied spatial statistics for public health data*. John Wiley and Sons, New Jersey.
- Wartenberg D, 1985. Multivariate spatial correlation: a method for exploratory geographical analysis. *Geogr Anal* 17:263-83.
- WHO, 2004. ICD-10: International Statistical Classification of Diseases and Related Health Problems, Tenth revision. World Health Organization, Geneva, Switzerland.
- WHO, 2015. Non-communicable diseases. Health in 2015: from MDGs to SDGs. World Health Organization, Geneva, Switzerland.
- WHO, 2018. World Health Statistics 2018: monitoring health for the SDGs, sustainable development goals. World Health Organization, Geneva, Switzerland.